用于 Hadoop 平台的混沌加密研究与实现 *

谢国波，姚灼琛

(广东工业大学 计算机学院，广州 510006)

摘　要：为解决传统单机模式串行加密方法存在的不足，设计了一种基于 Hadoop 平台的混沌加密算法的运行方案。该方案运用 MapReduce 并行框架和混沌加密伪随机数以及初值敏感的原理，提出一种针对 MapReduce 框架和混沌加密优化的并行混沌加密方案，即用明文长度作为初值，分别对 Chen、Lorenz、Rossler 三种超混沌系统进行初始迭代，同时提出对明文数据按 1 Mb 进行分块的设计理念，通过根据偏移量，判断每个分块生成长度为 1Mb 的 Chen、Lorenz、Rossler 三个密钥序列的方法，达到提升数据密度安全性、减少运行内存占有量等目的。该设计框架中，Chen 序列用于明文置乱操作，Lorenz 序列用于异或的扩散操作，Rossler 序列用于取模的辅助扩散操作。实验证明，针对 MapReduce 并行框架特性和混沌系统特性的优化算法，在能够有效减小内存占用量、又可以提高加密速度的同时，明文关联的加密操作达到了有效防御选择明文攻击的目的。

关键词：MapReduce；Chen；Lorenz；Rossler；超混沌系统；并行加密；

中图分类号：TP309.7　　doi: 10.3969/j.issn.1001-3695.2018.05.0351

Chaotic encryption research and application based on Hadoop

Xieguobo, Yaozhuochen

(*Faculty of Computer Guangdong University of Technology*, *Guangzhou* 510006, *China*)

Abstract: In order to solve the shortcomings of the traditional serial encryption method in single machine mode, such as the low density of data security, the difficulty in encrypting-efficiency to meet the requirement of increasing information data encryption and the excessive amount of memory in the running process, and so on. This paper designed a scheme of chaotic encryption algorithm based on Hadoop platform. The scheme used MapReduce parallel framework, chaotic encryption pseudorandom number and the principle of initial value sensitivity. A parallel chaotic encryption scheme for MapReduce framework and chaotic encryption optimization is proposed. That is, using the length of the plaintext as the initial value, uses Chen, Lorenz, Rossler chaotic system to generate secret key. At the same time, A the idea of dividing plaintext data into blocks according to 1Mb was proposed. According to the offset, each block can be judged to generate three key sequences of Chenn Lorenzl Rossler whose length is 1Mb. To improve the security of data density, reduce the amount of running memory and other purposes. In this design framework, Chen sequence is used for plaintext messing operation, Lorenz sequence is used for XOR diffusion operation and Rossler sequence is used for modular auxiliary diffusion operation. Experiments have proved that, for MapReduce, The optimization algorithm of using parallel framework and chaotic system can effectively reduce the amount of memory and improve the encryption speed, and the encryption operation associated with plaintext achieves the purpose of effectively defending against the attack of selected plaintext.

Key words: MapReduce; Chen; Lorenz; Rossler; hyper-chaotic system; patallel enctyption

0　引言

近年来，随着计算机网络及应用技术的迅猛发展，在信息领域所产生的数据量以指数级膨胀。随着企业、个人等隐私信息存在泄密或受到安全威胁的情况出现，人们对数据安全的需求也日益提高。近几年大数据网络安全问题已成学术界的热门讨论话题，这也成了信息安全领域有待开发提高的热点课题之一。面对网络及应用技术的升级换代，海量信息数据的产生，当前传统的单机模式串行加密方法因受其内存、处理器规模等硬性条件所限，在效率上和内存占用上已难以满足对大量数据

收稿日期：2018-05-25；**修回日期：**2018-07-18　　**基金项目：**广东省重大科技专项资助项目（2016B030306004）；广州市科技计划资助项目（201605101034176）

作者简介：谢国波（1977-），男，广东梅州人，教授，硕导，博士，主要研究方向为网络与分布式系统、复杂系统的分析与建模应用、基于 DSP 和 FPGA 平台的混沌信号产生与保密通信技术、信息网络研究（xiegb@gdut.edu.cn）；姚灼琛（1992-），男，硕士研究生，主要研究方向为混沌理论、大数据与云计算、机器学习．

的加密要求。而对于更大数量级的数据环境中，则面临计算负载可能远远超出单台计算机的运算极限等情况的困扰。针对上述弊端，提升数据安全保护变得愈发重要。随着云计算的出现为大数据加密提供了全新的思路。目前通用方法为建立云计算集群平台，通过将每个计算机封装成一个计算节点，每个节点可用以储存或处理海量数据。该法也因建成成本低而成为众多企业、科研单位等的合理选择。但是，该平台只为本文提供了有效处理海量数据的平台，却本身缺乏高效加密的手段，由此导致重要的科研数据、业务资料或客户材料等隐私信息存在容易泄密的危险。如何在大数据平台的基础上,设计一种针对 MapReduce 框架有所优化并行加密的算法，是本文要解决的问题。

本文所提用于 Hadoop 平台的混沌，是确定性系统中的随机运动。它的典型特征主要体现在两点：对初始条件的高度敏感性和良好的伪随机性质。它与传统密码学中的扩散、密钥、迭代等有许多相似之处。目前并没有太多关于混沌加密理论应用在 Hadoop 大数据平台的研究。2015 年，文献[5]提出了基于 MapReduce 的并行多混沌加密方案，利用 Logistic，Henon，Lorenz，Chen 等低维的混动系统，产生初值和干扰值对明文的对应 ASCII 码进行加密运算。2017 年，文献[6]提出了一种基于 Hadoop 大数据平台和无简并高维离散超混沌系统的加密算法，该法采用流密码对称加密方式,利用李氏指数均为正数的特点，设计高维离散超混沌系统进行加密，其优点体现在具有更好的统计特性。同年，文献[7]提出了在 Hadoop 平台使用 ARIA 算法的架构，该架构下用户可以选择 AES 或 ARIA 算法进行加密。基于上述文献阐述的算法，本文作者认为还有以下有待改进的地方：低维度的混沌系统的正李氏指数并不够大，没有很好的统计学特性产生的问题；并没有针对 MapReduce 并行计算框架进行优化。对每个分块的明文运算中出现太多冗余计算，因而占用了系统内存，减慢了运算速度。

为解决和优化上述不足，本文设计并实现了一种加密方案：即基于 Hadoop 的多个超混沌系统的算法，该算法首先，对明文数据按 1Mb 分组的同时根据明文数据的大小取出四个初值用于初始迭代，使之加密序列不受系统过渡过程的影响，同时又因混沌系统的伪随机性，初值和迭代次数相同的运算总会得出同样的密码序列的原理，故可在 MapReduce 开始阶段进行初始迭代，避免并行加密时的冗余运算。在每个 Mapper 运行时，根据偏移量依次生成 Chen、Lorenz、Rossler 的超混沌系统密钥，其中 Lorenz 系统的初值是 Chen 系统的 262144 次迭代值，Rossler 系统同理。对每一个 Mapper 读取的明文数据，根据其偏移量进行 Chen 密钥的置乱操作，Lorenz 密钥的异或操作，Rossler 密钥的取模操作。最后由 Ruducer 完成密文的拼接工作。本文设计的加密算法，不仅针对 MapReduce 框架作出优化处理，而且根据混沌系统自身的特性，有效的提高了加密速度，降低了内存占用量。

1　背景知识

1.1　混沌系统

1.1.1 Chen 超混沌映射

Chen 的超混沌系统[10]可以描述为

其中：a、b、c、d 为系统参数，当参数为 a=35,b=3,c=12,d=7 时，混沌系统的解如图 1 所示。

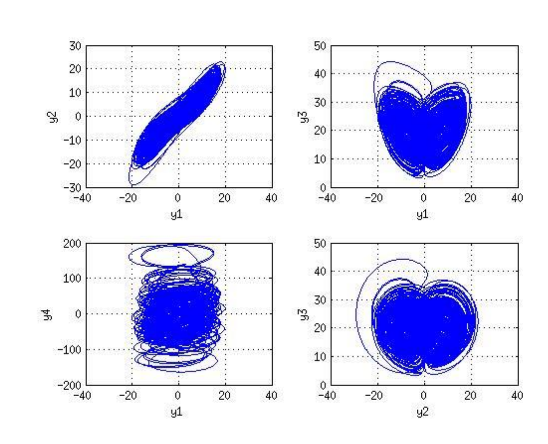


图 1　混沌系统 1

1.1.2 Lorenz 超混沌映射

Lorenz 的超混沌系统[1]可以描述为

$$\begin{cases} x = a(y - x) + w \\ y = cx - y - xz \\ z = xy - bz \\ w = -yz + rw \end{cases}$$

其中：a、b、c、r 为系统参数，当 a=10,b=8/3,c=28,r=-1 时，系统进入混沌，该参数下的混沌系统如图 2 所示。

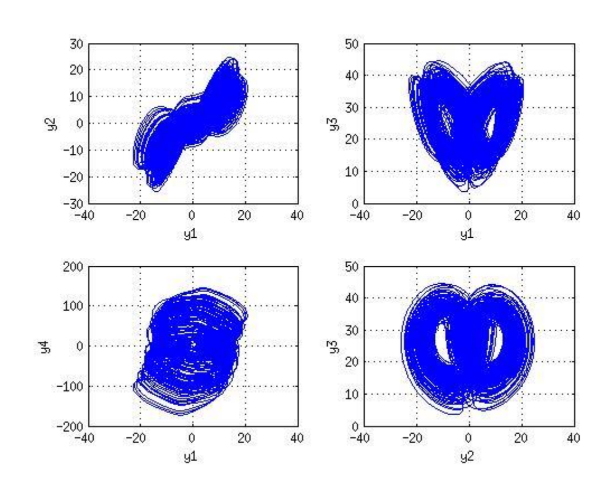


图 2 混沌系统 2

1.1.3 Rossler 超混沌映射

Rossler 的超混沌系统[10]可以描述为

其中：a、b、c、d 为系统参数，当 a=0.25,b=3,c=0.5.d=0.05 时，系统进入混沌态.

2　MapReduce 并行多混沌加密方案

本文所设计的这种基于 Reduce 的并行多混沌的超混沌加密方案，具有针对 MapReduce 框架优化，多混沌更多的正值李氏指数等特点，相较一般的加密算法有密钥空间大、安全性高、内存占用量更小、加密速度更快的优点。

2.1　具体加密方案

本文的多混沌加密方案运用了 Chen、Lorenz、Rossler 三个超混沌系统，通过置乱，正反异或操作、正反取模操作对明文进行加密。多个超混沌系统不仅各自的正李氏指数更多，而且相比单个混沌序列对大数据量的明文加密，其所需要的迭代次数也会更少，而单个混沌序列所需要的内存占用随着明文的数据量增大而增大，加密效率也随着数据量的增大而减少，显然不能很好的服务于大数据平台的海量数据源。将明文分块，并按照偏移量生成密钥序列，根据读入的明文偏移量同时进行加密，能够很好的解决该问题。具体的加密方案如下所示：

a)将明文按照每一块大小为 1Mb 分成 N 块。明文前 N-1 块大小为 1Mb，第 N 块为剩余数据。同时根据明文的大小，分别取千、百、十、个位数作为混沌系统的输入进行 300 次迭代跳过过渡态。

b)对于每一个分块，分别通过 Chen、Lorenz、Rossler 混沌映射生成密钥序列。Chen、Lorenz、Rossler 超混沌系统均为四维连续超混沌系统，利用欧拉法解出混沌系统，每个混沌系统每经过一次迭代则生成四个随机序列，即为十二个混沌序列：

其中:每个序列迭代次数为 262144 次（1/4 MB），Lorenz 系统的输入为 Chen 系统的第 262144 次迭代输出，Rossler 系统的输入为 Lorenz 系统的第 262144 次迭代输出。

c)对于生成的十二个混沌序列，仍然有较大的不足，比如均匀分布特性不佳；局部取值存在单调性；序列之间的相关度较高等等，所以在对明文进行加密之前仍需要分别对三个超混沌系统进行预处理，通过预处理去除各实数值的整数部分，统一值域；去除实数值后再将小数点后移 9 位，达到增强序列取值的无规则性和整体分布的均匀性[5]，具体操作如下式：

(1)

完成序列的预处理之后，根据 Map Reduce 的数据读取逻辑，取每一行明文数据对应的 ASC II 码 M，根据其偏移量选择对应的序列进行置乱，然后进行正反向异或扩散，最后进行正反向取模扩散操作进行二次干扰加密，算法运算流程图如图 3 所示，具体运算操作如下：

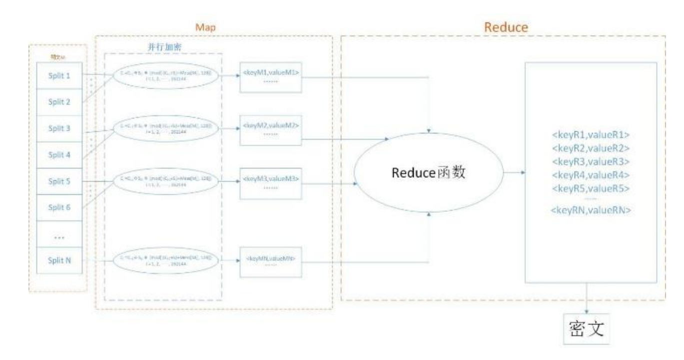


图 3　算法流程图

a)置乱操作。

(2)

b)正反向异或操作。

(a)正向异或。

(3)

(b)反向异或。

(4)

c)正反向取模操作。

（a）正向取模。

(5)

(b)反向取模。

(6)

经过上述加密操作后得到最终密文。

2.2　MapReduce 实现加密方案

MapReduce 并行框架中，主要由 Mapper 和 Reducer 组成，Mapper 主要作为父类，通过继承 Mapper 重写 map 方法实现对数据的逐行读取，进行相关的数据处理。map 方法将数据分成

了多个子任务，分配给多个处理器同时完成计算，Mapper 数量由分块数决定，每一个 Mapper 完成子任务的加密操作。最后交由 Reducer 负责整合所有 Mapper 输出的结果。综上可知，本文提出的并行加密方案中，每一块加密计算都是独立完成的，块与块之间并没有产生相互的干扰。因此 MapReduce 实现可以分为三个部分：Split，Map，Reduce。

2.2.1 Split

本文选择将输入数据按 1Mb 大小进行分割，这样一个 N Mb 的文件就会被分成 N 块。

算法 1 Split 算法

输入：加密文件

输出：1Mb 大小的文件分块

```
InputFormat()
isSplitable()
getSplits()
if((length!=0)&&isSplitable())
blocksize=1048576
return splits
```

那么对于前 N-1 个分块，分块大小都为 1Mb，并以 K-V 键值对的形式传输给 N 个 Mapper，每个 K（Key）为明文块的第一个字节偏移量，V（Value）为明文块的具体值。这样每个 Mapper 在进行加密处理的时候都是并行且无不干扰的。

2.2.2 Mapper

经过分块之后，每一个 Mapper 将单独计算明文的一部分。Mapper 之间的计算没有通信。每一个 Mapper 都进行 3.1 中的 B）至 D）运算。当偏移量==0 或 1048576-（mod1048576）<n 时，进行一次 B）运算，然后根据读入数据的每一行的偏移值找到对应的密钥序列，分别进行置乱，正反异或，正反取模的加密处理。最后输出为 K-V 对，K 为偏移量，V 为已加密的局部密文。

算法 2 Mapper

输入：<LongWritable,Text>

输出：<LongWriatble,Text>

```
if(==0&&1048576-（mod1048576）<n)
ChenGen()
LorGen()
RossGen()
=MessUp(,);
=Xor(,);
=Mod(,);
```

2.2.3 Reducer

当所有的 Mapper 完成了局部密文的加密操作，Reducer 将根据 Mapper 输入的 K 进行排序，排序后的密文即为完整的密文，Reducer 输出后便是所求密文。

2.3 算法安全性分析

2.3.1 密钥空间

基于 MapReduce 的并行多个超混沌加密方案，由于超混沌系统本身就具有很好的安全性，辅助 MR 每个 Mapper 单独计算的工作原理，使得这种加密方案不仅能够胜任海量数据的加密要求，并对超混沌系统的安全性没有消极影响。通过统计得出，该加密方案的密钥包括三个超混沌系统的初始迭代数 ，

，　；三个超混沌系统明文关联的初值

和参数

。

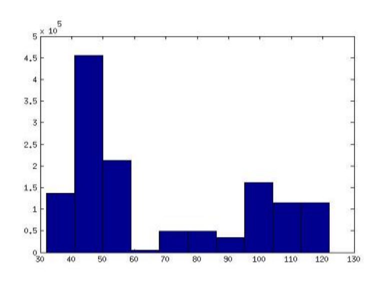
假设他们的值为：

初始迭代数 ，初值（1，1，1，1，1，1，1，1，1，1，1，1），以及各自超混沌系统的参数

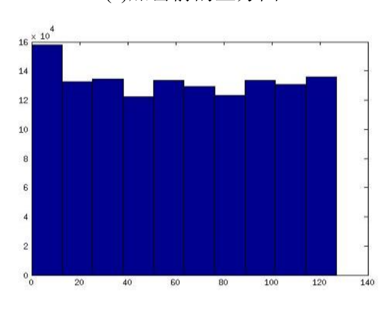
。则需要 10700 Byte 空间存放，即 85600 位，故破解组合数为 。可见密钥空间足够的大，足以抵抗暴力破解密钥攻击，并且随着初值的增大，密钥空间的增长也是指数级的。

2.3.2 直方图分析

另采用含有多个超混沌的加密算法的加密文本统计如图 4 所示。由统计直方图可知，明文经过加密算法的加密后，其统计特性改变相当明显。



(a)加密前的直方图



(b)加密后的直方图

图 4　统计直方图

2.3.3 初值敏感性

同时该算法的初值敏感性也非常良好，由于文本数据并不如图像数据在观察算法初值敏感性上直观，所以采用加密后的密文相关性比较，分别用 1 和 1.00000001 对同一个文件加密，

得出两次密文如图 5 所示，两次密文的相关性仅为 0.05130069。

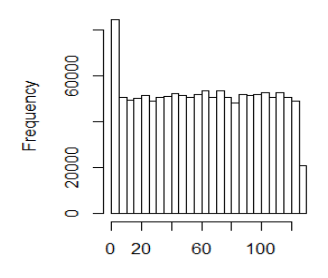


图 5　初值为 1 的密文的直方图

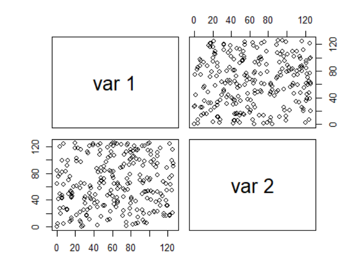


图 6.初值为 1.00000001 密文与初值为 1 的密文相关性比较

2.3.4 相关性分析

在 MapReduce 并行框架下对明文进行混沌加密算法与传统单机模式下的算法有根本的结构区别，因此考察算法的各个分块之间的互相关性是研究算法安全性的重要一环。令各个分块的明文数据的内容完全相同（图 7），对明文加密后各分块相关性如图 8 所示。结果表明，密文之间的相关性很小。

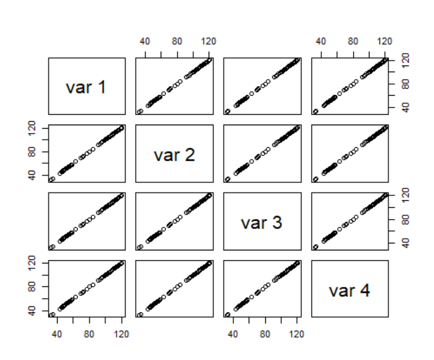


图 7　内容完全相同的明文

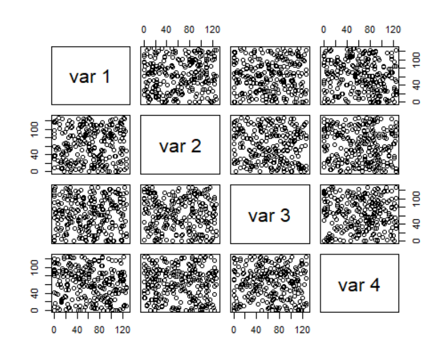


图 8　加密后各个密文的相关性

2.3.5 已知明文攻击与选择明文攻击

在加密过程中，密码系统的密钥参数 $C_{key}=(\chi_i,A_i,B_i,C_i,R_i),i=C,L,R$ 在每一次加密过程都保持不变，另外根据已知明文攻击的条件，攻击方获知任意明文，并且知道这些任意明文对应的密文。同理，选择明文攻击的条件中，攻击者通过选择对破译有利的明文以及对应的密文。

那么对于与明文无关联的算法，对于任意一次明文 $M(i),i=1,\ldots,N$ 加密，得到的对应加密密钥序列均为 $K_e=\{K_e(0),K_e(1),\ldots\ldots K_e(N)\}$，那么攻击方只需要两步便可破译密码。

获取等价密钥 $K_e(k)$

假设攻击方第一次进行加密的明文为 $M(1)$，则对应密文应为

$$C_1=C_{i-1}\oplus S_{ci}\oplus\{\mathrm{mod}[(C_{i-1}+S_i)+Mess[M_i],128]\}\ i=1,2\ldots262144$$

则攻击方可通过加密公式的逆运算获得对应 S_1，若不采取明文关联的密钥生成方案，则 $S_1=K_e(k)$

用等价密钥 $K_e(k)$ 破译加密后的明文

假设攻击者第 n 次加密的明文为 $M(n)$，而因等价密钥仍保持不变，故可获得对应密文 C_n，进而可以利用等价密钥破译相应的明文。

而本文设计的加密方案根据明文首字母的 ASCII 码作为密钥输入，参与到混沌序列的运算中，使得在加密不同的明文 $M(i),i=1,\ldots,N$，密钥 $K_e(k)$ 总是变化的，不再是保持恒定不变。由此可见，本文通过采用了与明文关联的方法，具有抵御已知明文攻击和选择明文攻击的能力。

3　实验结果以及分析

3.1　实验环境

采用的 Hadoop 大数据平台为 Master-Slaves 结构，由 4 个节点组成，包括 1 个 master 节点 和 3 个 slaves 节点。节点硬件配置为 CPU Inter i3-8100，3.60 GHz/6MB Cache，内存 24 GB，磁盘 256 GB SSD，千兆以太网卡。

操作系统为 Linux Ubuntu16.04，Hadoop 版本 2.6.5，Java 版本 jdk1.7.0_79，开发环境 Eclipse Luna Service Release 2（4.4.2）。

3.2　执行效率分析

在单个计算机运行加密算法的情况下，运行较大的密文加密运算时往往会受限于单机计算性能的局限，内存占用和运算速度都会极大的受到影响，甚至会出现内存溢出加密失败的问题。而基于 MapReduce 框架下的混沌加密算法能够很好的解决这个问题。运用本文设计的加密算法在加密 64 MB、128 MB、256 MB 的文本文件效率对比和内存占用量如图 9、10，表 7、8 所示。结果表明，针对 MapReduce 架构优化后的混沌加密算法相较单机模式以及 MapReduce 默认按行读取方式加密有更小的内存占用量和更快的加密速度。

单机环境下对 128 MB 大小的明文进行加密时就收到了内存的限制因内存溢出而出错，而加载 64MB 大小的文件情况也非常不乐观。而优化后的 MapReduce 并行算法通过每 1Mb 迭代生成长度为 1Mb 的密钥序列，有效的减少了生成混沌密钥时所需要的冗余迭代，同时因为加密完成后密钥序列便可以清空，

从而大大减少了运行时的内存占用量，可以看到在 Reduce 阶段，因为只用了一个核负责所有密文的拼接工作，所以 Reduce 在相同文件大小的情况下所需时间大致相同。

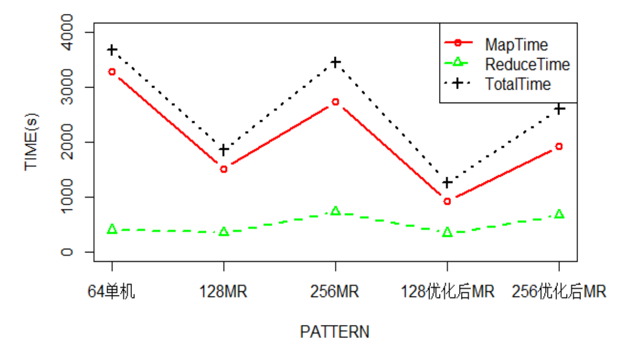


图 9　各个模式下的对应文件大小加密时间

表 7　各个模式的时间统计

时间/s	64 单机	128MR	256MR	128 优化后	256 优化后
Map	3273	1509	2725	931	1929
Reduce	402	348	728	339	672
Total	3675	1857	3453	1270	2601

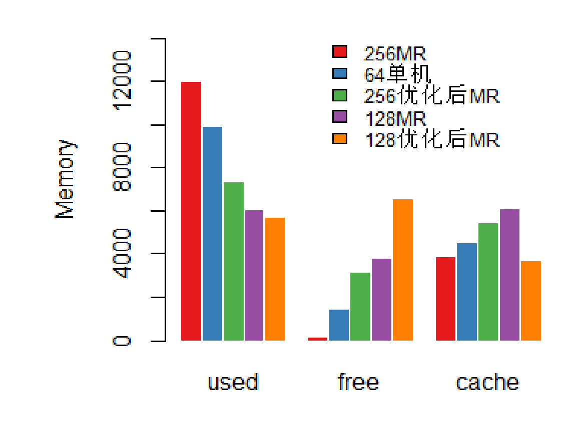


图 10　各个模式下的对应文件大小内存占用量

表 8　各个模式的内存统计

内存/MB	64 单机	128MR	256MR	128 优化后	256 优化后
map	9940	6086	11982	5737	7370
reduce	1504	3826	158	6582	3194
total	4573	6105	3876	3698	5453

4　结束语

本文提出了一种针对 MapReduce 和混沌系统优化的并行加密算法，并在 MapReduce 并行运算框架下实现了该算法，取得了预测的实验结果，结果表明本文中的算法能够实现海量数据的加密工作，有效地降低了加密时间，提高了加密效率，解决了加密运行时的内存占用量。本算法的创新点在于使用多个超混沌系统，利用其多个正李氏指数的良好统计学特性对密文加密，同时根据 MapReduce 并行框架的特点，修改了 MapReduce 默认的每个 Mapper 按行读取数据只能按行加密的方案，通过科学地分配 Mapper 处理明文分块，并在每个分块生成相应长度的密钥序列，在减少了冗余迭代的同时实现了内存的有效再分配。实验表明本算法在同类加密方案中提升了密文的安全性、同时具有更优秀的加密速度、减少了加密运行时的内存占用量，提供了更理想的内存空间。

参考文献：

[1] 张勇. 混沌数字图像加密 [M]. 北京: 清华大学出版社, 2016. (Zhang Yong, Chaotic digital image cryptosystem [M]. BeiJing: Tsinghua Universitiy Press, 2016.)

[2] Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters [J]. Communications of the ACM, 2008, 51 (1): 107-113.

[3] 孙燮华. 图像加密算法与实践: 基于 C#语言实现 [M]. 北京: 科学出版社, 2013. (Sun Xie Hua. Image encryption algorithms and practices with implementations in C# [M]. Beijing: Science Press, 2013.

[4] 肖锋, 张丽丽, 冯飞. 混合混沌系统的并行多通道彩色图像加密 [J]. 微电子学与计算机, 2016, 33 (8): 76-81. (Xiao Feng, Zhang Lili, Feng Fei. Parallel multi-channel color image encryption based on hybrid chaotic system [J]. Microelectronics and Computer, 2016, 33 (8): 76-81.

[5] 王欣宇, 杨庚, 闵兆娥 基于 MapReduce 的并行混合混沌加密方案 [J]. 计算机应用研究 2015, 32 (6): 1757-1760. (Wang Xingyu, Yang Geng, Min Zhaoe. Parallel mixed chaotic encryption scheme on MapReduce [J]. Application Research of Computers 2015, 32 (6): 1757-1760.)

[6] 温贺平, 禹思敏, 吕金虎. 基于 Hadoop 大数据平台和无简并高维离散超混沌系统的加密算法 [J]. 物理学报, 2017, 66 (23): 76-89. (Wen Heping, Yu Simin, Lyu Jinhu. Encryption algorithm based on Hadoop and non-degenerate high-dimensional discrete hyperchaotic system. Physics Letters A, 2017, 66 (23): 76-89.)

[7] Song Youngho. Design and implementation of HDFS data encryption scheme using ARIA algorithm on Hadoop [C]// Proc of IEEE International Conference on Big Data & Smart Computing. 2017: 13-16.

[8] Wang Xizhong, Chen Deyun. A parallel encryption algorithm based on piecewise linear chaotic map [J]. Mathematical Problems in Engineering, 2013, 2013 (5): 497-504.

[9] 蔡国梁, 黄娟娟. 超混沌 Chen 系统和超混沌 Rssler 系统的异结构同步 [J]. 物理学报, 2006, 55 (8): 3997-4004. (Cai GuoLiang, Huang JuanJuan. Synchronization for hyperchaotic Chen system and hyperchaotic Rössler system with different structure [J]. Physics Letters A, 2006, 55 (8): 3997-4004.

[10] Wang Xingyuan, Zhangm Yingqian, Bao Xuemei. A novel chaotic image encryption scheme using DNA sequence operations [J]. Optics and Lasers in Engineering, 2015, 2015 (73): 53-61.

[11] Wang Xingyuan, Liu Lintao, Zhang Yingqian, A novel chaotic block image encryption algorithm based on dynamic random growth technique [J]. Optics and Lasers in Engineering, 2015, 2015 (66): 10-18.

[12] Zhang Yingqian, Wang Xingyuan, A new image encryption algorithm based

on non-adjacent coupled map lattices [J]. Applied Soft Computing, 2015, 2015 (26): 10-20.

[13] Wang Xingyuan, Feng Chen, Tian Wang. A new compound mode of confusion and diffusion for block encryption of image based on chaos [J]. Communications in Nonlinear Science and Numerical Simulation, 2010, 15 (9): 2479-2485.

[14] Liu Hongjun, Wang Xingyuan. Color image encryption based on onetime keys and robust chaotic maps [J]. Computers & Mathematics with Applications, 2010, 59 (10): 3320-3327.